Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes

Mueller, M F ; Kegel, A ; Schimmel, S M ; Dillier, N ; Hofbauer, M

Abstract: Sound localization with hearing aids has traditionally been investigated in artificial laboratory settings. These settings are not representative of environments in which hearing aids are used. With individual Head-Related Transfer Functions (HRTFs) and room simulations, realistic environments can be reproduced and the performance of hearing aid algorithms can be evaluated. In this study, four different environments with background noise have been implemented in which listeners had to localize different sound sources. The HRTFs were measured inside the ear canals of the test subjects and by the microphones of Behind-The-Ear (BTEs) hearing aids. In the first experiment the system for virtual acoustics was evaluated by comparing perceptual sound localization results for the four scenes in a real room with a simulated one. In the second experiment, sound localization with three BTE algorithms, an omnidirectional microphone, a monaural cardioid-shaped beamformer and a monaural noise canceler, was examined. The results showed that the system for generating virtual environments is a reliable tool to evaluate sound localization with hearing aids. With BTE hearing aids localization performance decreased and the number of front-back confusions was at chance level. The beamformer, due to its directivity characteristics, allowed the listener to resolve the front-back ambiguity.

DOI: https://doi.org/10.1121/1.4705292

Originally published at:
Mueller, M F; Kegel, A; Schimmel, S M; Dillier, N; Hofbauer, M (2012). Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes. Journal of the Acoustical Society of America, 131(6):4732-4742.
DOI: https://doi.org/10.1121/1.4705292

Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes

Martin F. Mueller,[a] Andrea Kegel, Steven M. Schimmel, and Norbert Dillier
Laboratory for Experimental Audiology, University Hospital Zurich, Zurich 8091, Switzerland

Markus Hofbauer
Phonak AG, Laubisrütistrasse 28, 8712 Stäfa, Switzerland

(Received 4 January 2011; revised 4 March 2012; accepted 12 March 2012)

Sound localization with hearing aids has traditionally been investigated in artificial laboratory settings. These settings are not representative of environments in which hearing aids are used. With individual Head-Related Transfer Functions (HRTFs) and room simulations, realistic environments can be reproduced and the performance of hearing aid algorithms can be evaluated. In this study, four different environments with background noise have been implemented in which listeners had to localize different sound sources. The HRTFs were measured inside the ear canals of the test subjects and by the microphones of Behind-The-Ear (BTEs) hearing aids. In the first experiment the system for virtual acoustics was evaluated by comparing perceptual sound localization results for the four scenes in a real room with a simulated one. In the second experiment, sound localization with three BTE algorithms, an omnidirectional microphone, a monaural cardioid-shaped beamformer and a monaural noise canceler, was examined. The results showed that the system for generating virtual environments is a reliable tool to evaluate sound localization with hearing aids. With BTE hearing aids localization performance decreased and the number of front-back confusions was at chance level. The beamformer, due to its directivity characteristics, allowed the listener to resolve the front-back ambiguity.
 [http://dx.doi.org/10.1121/1.4705292]

PACS number(s): 43.66.Qp, 43.66.Ts [TD] Pages: 4732–4742

I. INTRODUCTION

The human auditory system is constantly engaged in the identification and localization of various competing sources in complex acoustical environments. The everyday soundfield typically contains background noise, reverberance and simultaneous sound events coming from different directions. Despite the complexity of the acoustical scenes, the binaural auditory system is able to effectively separate and localize sound sources of interest. Sound localization is affected by background noise, reverberation and interfering signals among others (Good and Gilkey, 1996; Lorenzi *et al.*, 1999; Langendijk *et al.*, 2001). To localize sound sources the human auditory system uses mainly interaural time and level differences (ITDs and ILDs). Additionally, the spectral filtering induced by the pinna allows the identification of the elevation of the sound sources. Pinna cues are also essential to resolve front-back confusions.

Sound localization with bilateral hearing aids has been investigated in various recent studies with different device types, listening configurations, algorithms and microphone positions. Questionnaire surveys indicated clear benefits in sound localization for patients fitted with bilateral hearing aids compared to unilateral fittings for every type of device (Boymans *et al.*, 2009; Noble and Gatehouse, 2006). Listening experiments carried out in the laboratory, however, indicate a degradation in localization performance caused by bilateral hearing aids compared to unaided conditions (Van den Bogaert *et al.*, 2011; Best *et al.*, 2010; Van den Bogaert *et al.*, 2006; Keidser *et al.*, 2006; Köbler and Rosenhall, 2002; Noble and Byrne, 1990). In these studies, when hearing-impaired listeners were tested, the signals in the unaided conditions were played at equal loudness levels. The results suggest that, while hearing impaired subjects benefit from the amplification provided from the second hearing aid, the signal processing in the devices distorts essential localization cues.

Several factors are detrimental for the localization of sound sources with bilateral hearing aids. Keidser *et al.* (2006) investigated the effect of multi-channel compression, noise reduction and directional microphones on horizontal sound localization. Their study included Behind-The-Ear (BTE), In-The-Ear (ITE) and Completely-In-the-Canal (CIC) hearing aids, considering thus microphone position effects as well. Their results showed that compression and noise reduction distorted ILDs, which led to a poorer performance. Furthermore, the position of the microphones of BTE hearing aids reduces pinna cues that are used to distinguish sounds from the front and the back. This has been confirmed in various studies (Van den Bogaert *et al.*, 2011; Best *et al.*, 2010; Keidser *et al.*, 2006; Köbler and Rosenhall, 2002). The use of directional microphones can reduce the number of front-back confusions (Keidser *et al.*, 2006).

The experiments reported previously have been carried out in the laboratory with different degrees of complexity but represent nevertheless artificial situations. In Keidser *et al.* (2006), for example, one test condition included the presence of a constant interfering noise at 80° of the listeners whereas the other algorithms were evaluated in

[a]Author to whom correspondence should be addressed. Electronic mail: marmuell@gmail.com

quiet. In Van den Bogaert *et al*. (2006), sound localization with bilateral hearing aids was evaluated in a moderately reverberant setting. In one condition, interfering multitalker babble noise was played at defined positions at the sides of the listeners. Hearing aids need to be evaluated in acoustical environments in which they are commonly used, because noise suppression algorithms affect auditory cues differently in noisy environments, depending on the type, the level and the position of the noise. Reflections might diminish the effectiveness of beamforming techniques as well.

Virtual acoustics can be used to evaluate hearing aid algorithms in more realistic environments. It is a relatively simple and convenient method for reproducing virtual spaces. This technique combines Head-Related Transfer Functions (HRTFs) and room simulations and theoretically allows the reproduction of any sound field at the eardrums of the listener (Moeller, 1992). The use of virtual acoustics enables the evaluation of existing hearing aid algorithms and research prototypes in the most diverse and relevant listening environments. The hearing aids can be implemented offline, which allows the evaluation of the most advanced algorithms. The realism of sounds generated with virtual acoustics and its impact on sound localization have been investigated in numerous studies. It has been shown that virtual sound sources can be localized as accurately as real sources when individual HRTFs are used (Bronkhorst, 1995; Wightman and Kistler, 1989).

In this study, four different scenes in diffuse background noise and three hearing aid algorithms were implemented. The scenes were generated using individual HRTFs and played via speakers located in the ear canals of the test subjects. The virtual environment was simulated using the ROOMSIM software (Schimmel *et al*., 2009). The simulator uses an image source model to simulate early reflections. This is combined with a stochastic process that models late reflections. A similar room simulation procedure was used by Rychtarikova *et al*. (2009). In their study, sound localization and speech intelligibility have been compared between a real and a simulated playback room. In the latter condition, the Binaural Room Impulse Responses (BRIRs) were generated using HRTFs measured on an artificial head. Their results show an increase in front-back confusions for the virtual condition. It is possible that the use of non-individualized HRTFs and the impossibility to make head movements partly increased the rate of errors. No change in speech intelligibility was noticed between the two reproduction methods.

In background noise or in the presence of competing interference, sound localization degrades with decreasing signal-to-noise ratio (SNR) (Lorenzi *et al*., 1999; Good and Gilkey, 1996) or when the interferer is located close to the target signal (Langendijk *et al*., 2001). In these conditions, front-back confusions and the perceived elevation of the source are most affected by the interference. Front-back confusions however can be resolved by head movements, as shown by Wightman and Kistler (1999) and Wallach (1940). Using slight head movements, the listener can resolve ambiguities in the horizontal cues and differentiate a sound in the back from the front and vice versa.

In the first experiment presented in this study, the virtual playback system was evaluated. The scenes were either played through a ring of loudspeakers located in a real room or reproduced virtually. Sound localization was then compared between the real and the virtual playback rooms. In the second experiment, the usefulness of the system for hearing aid testing was evaluated. Three standard BTE hearing aid algorithms were tested, namely, an omnidirectional microphone, a cardioid-shaped beamformer and a noise canceler.

II. METHOD

A. Reference conditions

Four different scenes were selected based on their everyday relevance. The experiment required the localization of four different test signals that favor different ranges of localization cues. The four scenes are:

(1) a man speaking in a crowded cafeteria;
(2) a phone ringing in a busy office;
(3) an ambulance siren on a busy street;
(4) a bird singing in a windy forest.

The corresponding spectrograms of the four target signals are shown in Fig. 1.

The room where the localization experiments were carried out was an acoustically treated shoebox-type room with octave-band reverberation times (T_{60}) shown in Table I. The room was 6.53 m long, 5.72 wide, and 2.34 high. The receiver was set at position (3.69, 2.85, 1.15) facing the long wall. The sounds were played through a loudspeaker ring centered on the receiver position at a distance of 1.5 m with an angular spacing of 30°.

The background noise consisted of single channel recordings of ambient sounds. The signals were sampled in 12 segments of 8 s each. The starting points of the segments were chosen randomly along the initial sound signal. The 12 signals were then played simultaneously over the loudspeaker ring, creating a diffuse soundfield around the listener. All recordings were done using omnidirectional microphones. Target and noise were recorded separately.

For the four scenes the SNR was set to 3 dB based on their rms values. This SNR was considered as containing sufficient noise for the hearing aid algorithms to work properly while maintaining good localization performance. The level of the background noise was set to 60 dB at the center of the loudspeaker ring.

In experiment I, three conditions were tested. In the first condition, the scenes were played through the loudspeaker ring in the real room. The test subjects listened with their "own ears." This is the absolute reference condition and is referred to as *ls_open*. The second condition (*sim*) evaluates the system for virtual acoustics. The playback room was simulated and the sound was played through small speakers located in the ear canals. The HRTFs used for the simulations were measured using the same devices. Due to their size, the ear canal speaker-microphone systems might modify monaural spectral cues and influence negatively the localization of sound sources. Therefore, the condition was included in which the scenes were played by the external

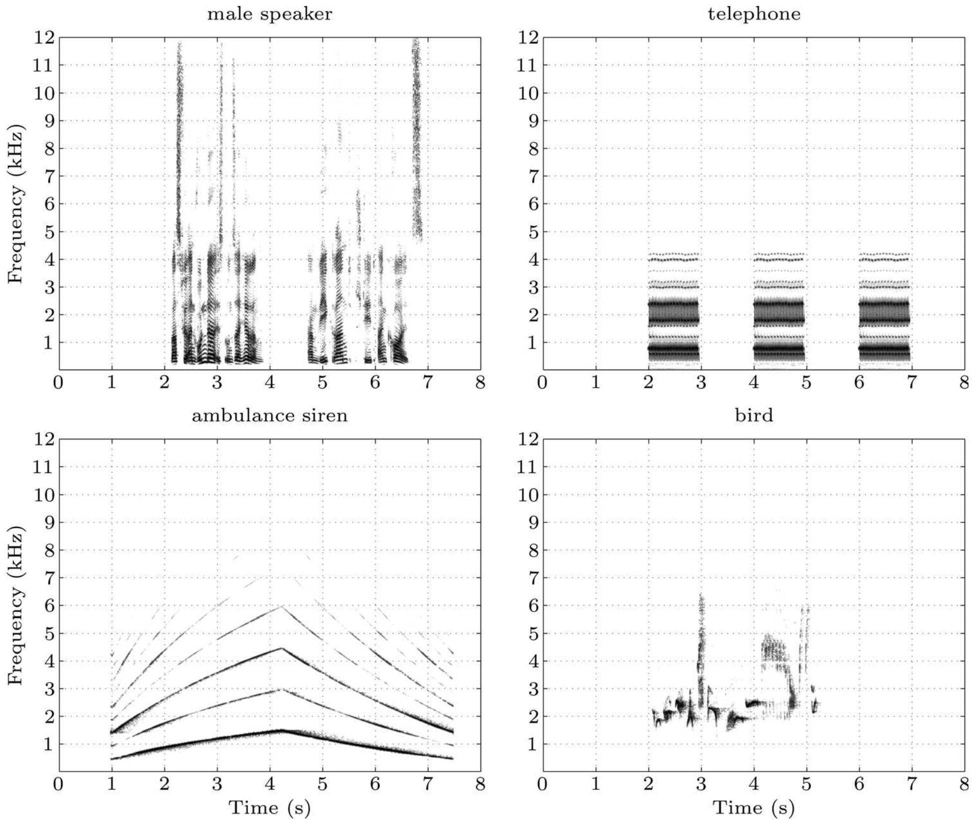


FIG. 1. Spectrograms of the target signals used in the localization experiment. Four scenes were implemented: a cafeteria, an office, a street and a forest. The listeners had to localize (top left) a male speaker, (top right) a phone, (bottom left) an ambulance siren, and (bottom right) a bird, respectively.

loudspeakers while the test subjects wore passive speakers (*ls_cic* condition). The ear canal transducers are described in details in Sec. II. In experiment II, the conditions tested are called *omni, beam* and *NC* for the omnidirectional, beamformer and noise canceler algorithms, respectively. A description of the algorithms is given in the following section.

B. Hearing aid algorithms

The first implemented algorithm was the omnidirectional microphone configuration. In this case, the scenes were simulated using the front microphones of the BTEs only. No processing was done by the hearing aids. This condition investigated the effect of the microphone position on sound localization.

The second algorithm was a first order differential static beamformer. It had a cardioid directional characteristic and reduced sound coming from 180°. The directivity pattern was obtained by delaying the signal of the rear microphone. The frequency-dependent phase shifts depended on the distance between the front and back microphones and on the individual HRTF characteristics. The differential processing of the algorithm introduced a highpass behavior. A lowpass filter compensated for this effect (Hamacher *et al.*, 2005).

TABLE I. Octave band reverberation times of the measured and the simulated rooms in [ms].

frequency [Hz]	125	250	500	1000	2000	4000	8000
$T_{60}meas$ [ms]	230	270	270	210	230	300	300
$T_{60}sim$ [ms]	229	271	273	213	229	304	331

The noise canceler was a Wiener filter type implementation. The incoming signal was divided into frequency bands. For each subband, the power spectra of the noise and of the speech were estimated. Subbands with high noise, i.e., low SNR, were attenuated whereas subbands with high SNR were unchanged. The SNR estimator was based on the assumption that the noise signal was relatively stationary, whereas the target was more heavily modulated (Hamacher *et al.*, 2005).

Since both monaural algorithms modify level and phase independently in each hearing aid on the left and right side, ITDs and ILDs will potentially be modified. The noise canceler, however, does not change ITDs. Both algorithms were implemented on a Simulink platform and all the processing was done offline, prior to the first test session.

C. Virtual sound reproduction

The sound recording and playback device consisted of a customarily designed pair of miniature microphone-speaker systems located inside a subject's ear canal. They were mounted on an open shell of CIC hearing aids. The devices were manufactured individually for every test subject prior to the experiment. The choice of the open CIC system over headphones was due to the following reasons: first, the ear canal was open during playback. This improved the reproduced spatial image and reduced the effect of sound internalization (Kim and Choi, 2005). Second, the system always stayed at the same location in the ear canals. The system therefore did not need to be calibrated at each utilization. Finally, being an open system, it allowed a direct comparison between loudspeaker and simulated playbacks.

1. HRTFs measurements

The HRTFs were measured in a low reverberant sound-treated room using the maximum-length sequence (MLS) technique (Rife and Vanderkooy, 1989). They were recorded using the microphone of the open CIC systems. Reflections were removed from the HRTFs by trimming the impulse responses 4 ms after the first peak. The MLS signals were played at 70 dB SPL. The sequence was sampled at 44.1 kHz and lasted 6 s. The recordings were done using the same loudspeaker arrangement as described in Sec. II A. The resolution of the HRTFs was thus 30°. HRTFs were measured for each participant at the beginning of the first test session.

To complete the set of measured positions, the recorded HRTFs were merged into a set of anechoic KEMAR HRTFs (Gardner and Martin, 1994). The KEMAR data set consists of HRTFs recorded on dummy head for 710 positions, ranging from elevation angle $-40°$ to $90°$ with a minimal azimuthal separation of 5°. The direct sound component of the simulated BRIRs was always composed of the individual recorded HRTFs. The KEMAR HRTFs were exclusively used for simulating reflections where no measured transfer function was available. The generation of the BRIRs is described in details in Sec. II C 4.

2. BTE HRTFs interpolation

The set of BTE Head-Related Transfer Functions (BRTFs) was recorded by a pair of standard BTE hearing aids each with two microphones at 12 mm distance. They were measured in the same room as the HRTFs and using the same procedure. The set of BRTFs was interpolated to a collection of transfer functions of the same format as the KEMAR HRTFs, covering the same positions. This was done because the algorithms are very sensitive to phase and amplitude differences between the BRTFs of the front and rear microphones. The combination of the BRTFs with unprocessed KEMAR data would reintroduce absent pinna cues as well.

The interpolation of BRTFs was carried out after time-alignment of the transfer functions. It has been shown that the performance of interpolation in the time or frequency domain can be improved by compensating HRTFs prior to interpolation according to the time of arrival of sound (Matsumoto *et al.*, 2004). That is, the HRTFs were time aligned and interpolation was carried out on the time-aligned HRTFs. In order to achieve sub-sample precision in the time alignment, the time of arrival itself was also interpolated. For positions in the horizontal plane, the BRTFs were linearly interpolated after time-alignment by a factor of 6 giving a resolution of 5°.

For the transfer functions corresponding to positions of different elevations, the delays to the front and back microphones were obtained using the spherical-head model described in Duda and Martens (1998). This procedure ensured that the delays between the front and back microphones are realistic. The amplitudes were obtained by interpolating the measured BRTFs at the corresponding azimuths in the horizontal plane. The interpolated BRTFs were used only for simulating reflections.

3. HRTF and BRTF calibration

The HRTFs and BRTFs were measured at different positions at the ears. This induced coloration differences that needed to be compensated before playback. The equalization of the transfer functions was done using the diffuse calibration method described by Moeller (1992, Sec. 5.2, p. 197). According to this technique, the transfer functions were averaged across all measured positions. The transfer functions were then divided by the average filter of the measured positions and multiplied by the average filter of the playback positions. This removed effectively the coloration differences between two transfer functions.

4. Room modeling and simulation

The virtual room was a simulation of the room described in Sec. II A. It was modeled with the ROOMSIM software. The surface absorption parameters of the ROOMSIM simulator were set to fit reverberation times measured in this playback room. The surface diffusity parameters of the ROOMSIM software were adjusted in order to match the level and the diffusity of the reflections and to minimize the perceptual differences between the simulated and measured impulse responses. The direct sound component of the BRIRs was composed of the individual recorded HRTFs or BRTFs. In this setup, most of the reflections were simulated using KEMAR HRTFs or interpolated BRTFs.

The directivity of the loudspeakers was modeled as a three-dimensional cardioid, pointing towards the receiver. This implied that most of the reflective energy came from the floor, the ceiling and the facing walls.

D. Test procedure

For each test condition, the test subject was asked to localize the target sound source (i.e., the male speaker, the phone, the ambulance siren, the bird) in the situation-specific background noise. Every test condition started with an orientation session, in which the scene was presented to the test subject. In this training round, all the 12 positions were played one after the other starting from the front and moving counter-clockwise. The subject could follow the position of the sources on a touch screen located in front of him. The diffuse background noise was played continuously. This was followed by a second training session, where every target position was presented once. Before the actual test run, the test subjects had to point out on the screen the position where they heard the sound coming from. Feedback was provided. Every position was presented twice in random order, resulting in 24 stimuli. The test subject had to indicate the position of the target source on the touch screen. No feedback was provided. The touch screen symbolically represented the test scene (i.e., cafeteria, office, street, forest) with 12 buttons arranged around a schematical listener. The subjects were instructed not to move their head during the experiment. A typical test run lasted approximately 10 min.

The 12 test conditions of the first experiment (4 scenes $\times$ 3 playback modes) were divided in three blocks of four. The eight conditions where the subjects wore the open CIC devices where randomly mixed. The four other conditions were presented in one block, in random order. This made it more comfortable for the test subject, as the open CICs did not need to be constantly inserted and removed between two successive tests. The three blocks were presented in random order. After one block was completed (approximately 40 min), the subjects took a break. The test subjects who completed the first experiment, returned on another day for the second experiment. The 12 test conditions (4 scenes $\times$ 3 hearing aid algorithms) were randomly mixed in three blocks of four. The second experiment followed the same test procedure as the first one.

At the beginning of the experiment, the test subject was asked to match the level of the simulation to the level of the loudspeaker presentation. To do this, the listener could switch between loudspeaker presentation and simulation to compare both loudness levels. He could increase and decrease the level of the simulation in steps of 1 dB until it matched the level of the external presentation.

E. Test subjects

Twelve normal-hearing subjects took part in the experiment (9 males, 3 females, age 35 $\pm$ 7 years). All subjects were confirmed to have hearing thresholds greater than -20 dB across all frequencies.

F. Data analysis

Here 0° was defined here as the position directly in front of the listener, 90° as the position to the left of the listener, and 270° as the position to the right of the listener. The localization performance was evaluated in two different ways. The accuracy of the directional localization was measured using the angular root-mean square (rms) error. As another indicator of the quality of the simulation, the amount of front-back confusions (fb) was considered. Front-back confusions occur when a sound presented in the front is heard in the back and vice versa. Those two phenomena represent different types of errors and were analyzed separately. Furthermore, the standard angular rms error is particularly sensitive to front-back confusions. Such confusions cause large errors for positions where the directional information was perceived and reported correctly. To remove this effect, the front-back confusions were resolved prior to measuring the directional error, which has commonly been done in localization experiments (Langendijk *et al.*, 2001). The angular rms error rms_θ is defined for each position as such:

$$\mathrm{rms}_\theta = \sqrt{\frac{\sum_{i=1}^{N}\left(\arcsin(\sin x_\theta) - \arcsin\left(\sin y_{\theta,i}\right)\right)^2}{N}}, \qquad (1)$$

where x_θ is the position played at angle θ and $y_{\theta,i}$ the response given by the test subject at test iteration i. N is the total number of repetition. Equation (1) implies that for a sound source played at 30°, 30° and 150° are considered to be correct answers. An average rms error taken over all played positions characterizes a subject's directional performance for a given test condition.

The amount of front-back confusions was evaluated as a percentage of occurrence over all possible confusions. Positions played at 90° and 270°, for which front-back confusions are not defined, were ignored. Sounds incorrectly located at 90° and 270° were not considered as confusions. These corrections result in a chance level of 41.66%.

III. RESULTS

A. Experiment I: Evaluation of the virtual acoustics system

Figure 2 shows the rms_θ averaged across all test subjects for every test condition. The rms_θ varies considerably across position and across scene. In all scenes except the cafeteria, the sound was accurately localized in the front, but rather poorly on the sides or in the back. The same pattern appears for presentation over loudspeaker with or without CICs (*ls_open* and *ls_cic* conditions in the upper and middle panels) and for the simulation (*sim*, in the bottom panel). The four scenes were not perceived as equally difficult. The male speaker was easily localized whereas the bird's position was frequently misjudged.

The upper panel of Fig. 3 shows the rms error, averaged across test subjects, for every test condition along with one standard deviation. The test subjects performed differently in the four different scenes. The overall results for each test condition are shown in Table II.

Significant differences between the rms error and the amount of front-back confusions for the four scenes and the three reproduction methods were examined using a one-way analysis of variance. Significance was set at $p < 0.05$. No significant difference in terms of rms error between the three

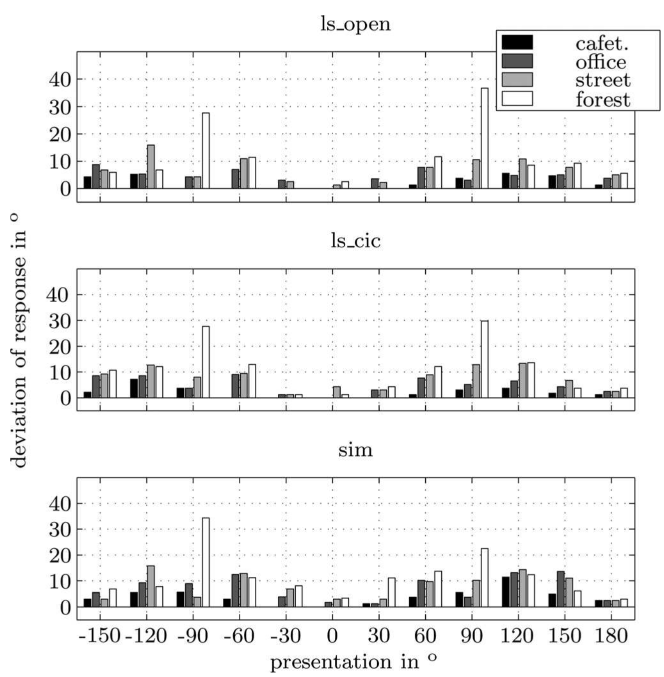


FIG. 2. Mean angular rms error, rms_θ, for the different scenes for the three sound reproduction methods. *ls_open* denotes loudspeaker playback with open ear canal (the natural listening condition), *ls_cic* stands for loudspeaker playback with the open CICs in the ears and *sim* represents the fully simulated environments with sound playback through the open CICs.

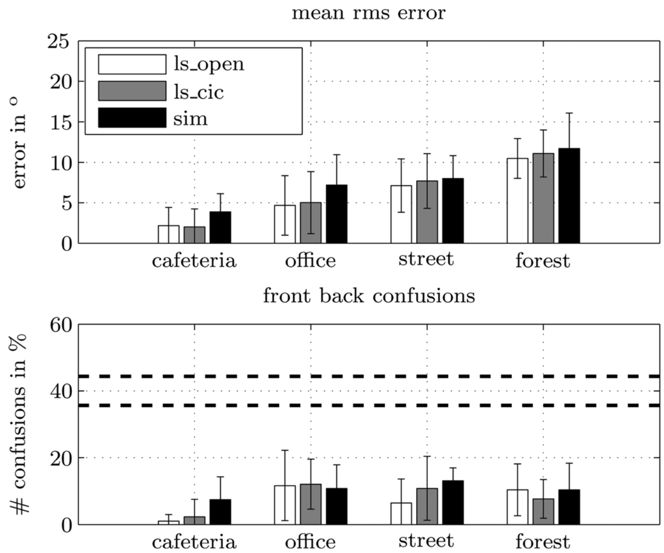


FIG. 3. Mean rms error (above) and percentage of front/back confusions (below) for each reproduction methods for the different scenes. The error bars show 1 standard deviation. Chance level, along with 95% confidence interval is plotted in dashed.

reproduction methods was found for the *office, street*, and *forest* scenes ($p > 0.11$).

The localization was only significantly worse in the simulated *cafeteria* condition (rms, fb: $p = 0.05$). The average rms error of the *sim* condition in this scene was, however, very small (3.9°). For the other sound reproduction methods, the localization of the target speaker was nearly perfect with a directional error of at most 2.2° and 2.3% of front-back confusions. The passive open CICs in the ear canal did not impair localization performance ($p \geq 0.23$).

The amount of front-back confusions varied greatly with the test subjects. As a result, the standard deviations were very large. The bottom panel of Fig. 3 shows the percentage of confusions for the different scenes and reproduction methods. The dashed lines show the chance level along with the 95% confidence interval. Results falling in this interval can be considered to follow with a 95% certainty a random guessing strategy.

For the *office* and *forest* scenes, the amount of front-back confusions was similar for the three reproduction methods. In these conditions, the simulations did not affect localization ability. In the *cafeteria* scene, the simulations were significantly worse than the *ls_open* and *ls_cic* conditions ($p \leq 0.05$). 7.5% of the signals were incorrectly localized in the front or in the back, which significantly more than 1.0% and 2.3% for the *ls_open* and *ls_cic* conditions respectively. In the cafeteria condition, the target signal was the most broadband of the stimuli presented, containing low frequencies components. At low frequencies, the human auditory system is sensitive to ITDs as small as 10 μs (Hershkowitz and Durlach, 1969). At a sampling rate of 44.1 kHz, this corresponds to half of the sample interval. Measuring ITDs with this precision is difficult. This would explain why the *cafeteria* case was the only scene where the *sim* condition yielded significantly worse performance in terms of directional and front-back errors, even though it was perceived as the most easy by the test sujects. Although the subjects were told to keep their head in a fixed position, unintentional head movements could have helped to resolve the front-back confusions when the sound was played through the loudspeakers. The virtual system was not designed to respond to head movements.

The four scenes were not perceived as equally difficult. This was desired, as the aim of the second experiment was to explore the weaknesses of different hearing aid algorithms. Scenes with different characteristics and degrees of difficulty permit to better rate and evaluate the hearing devices. The results clearly showed a change in rms error. Averaged across the three reproduction methods, the rms error was 2.1° for the *cafeteria*, 4.5° for the *office*, 5.9° for the *street* and 9.7° for the *forest* condition. The rms error differences between the *cafeteria* and *forest* scenes and the other tested environments were statistically significant ($p \leq 0.05$). The amount of front-back confusions was of the same order between the different scenes, with the *cafeteria* showing slightly fewer mistakes (3.6% vs 11.4%, 8.9% and 10.2% for the *office*, *street* and *forest*, respectively).

The rms error was higher in the back than in the front, as illustrated in Fig. 4. For signals played in the front, the statistical analysis showed again that performance in the *cafeteria* and *forest* scenes was significantly worse for the *sim* condition ($p \leq 0.05$).

The open CICs affect mostly the high frequency content of the signals, due to their small size. For signals played in the back, high frequencies are naturally attenuated by the pinna. An inaccurate reproduction of high frequencies has therefore less effect than for signals played from the front. This could be an explanation for the difference in localization performance

TABLE II. Mean results and standard deviations for all the scenes tested. The last column shows performance averaged across scenes. f-b denotes front-back confusions in percent.

	cafeteria			office			Street			forest			all		
	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim	ls_o	ls_c	sim
rms (°)															
mean	2.2	2.0	3.9	4.7	5.0	7.2	7.1	7.7	8.0	10.5	11.1	11.7	9.1	9.3	10.4
Std	2.3	2.2	2.2	3.7	3.8	3.7	3.3	3.4	2.8	2.5	2.9	4.4	2.5	2.6	2.7
f-b (%)															
mean	1.0	2.3	7.5	11.7	12.1	10.8	6.5	10.8	13.1	10.4	7.7	10.4	7.4	8.2	10.5
Std	2.0	5.3	6.8	10.5	7.4	7.1	7.2	9.6	3.9	7.7	5.8	8.0	5.1	6.0	4.2

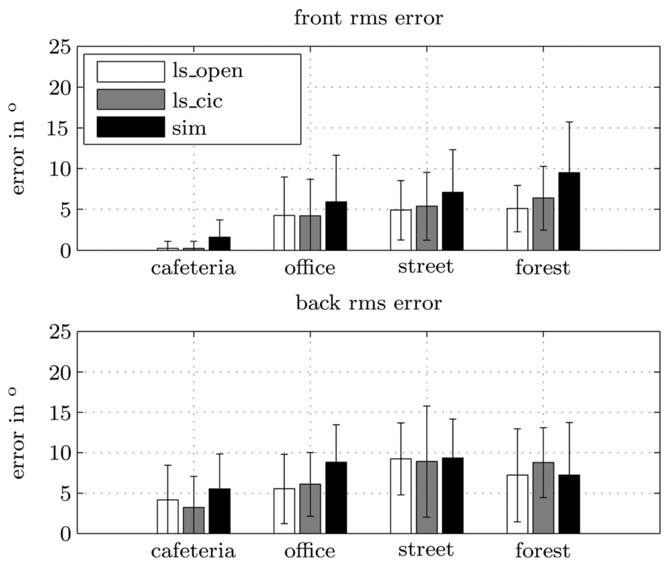


FIG. 4. Mean rms error for positions played at front ($|\theta| \leq 60^{\circ}$, above) and in the back ($|\theta| \geq 120^{\circ}$, below).

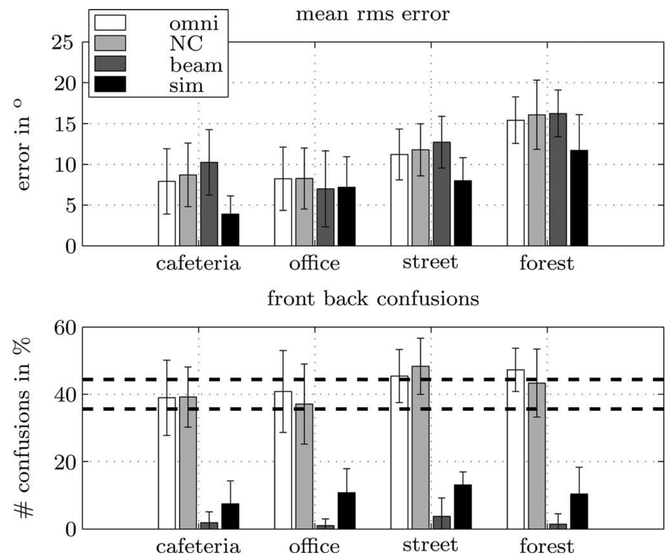


FIG. 5. Mean rms error (above) and percentage of front/back confusions (below) for the *sim* (reference, taken from Fig. 3), *omni*, *NC* and *beam* algorithms for the different scenes. The error bars show one standard deviation. Chance level along with 95% confidence interval lays between the two dashed line.

that can be seen in the front, but not in the back. No significant differences were found for back positions.

The effect of learning on the performance of the test subjects was further examined. No significant difference was found between the test and retest sessions.

B. Experiment II: Evaluation of BTEs algorithms

In the second experiment, the localization task was repeated with the same subjects and the same scenes. All the signals were processed offline and presented through the open CICs. Three different BTE algorithms were evaluated: the omnidirectional case (*omni*), where no processing was done by the hearing aid, the beamformer (*beam*) and the noise canceler (*NC*) conditions with active BTE processing. The results were analyzed in the same way as for experiment I. They are shown in Fig. 5, with the upper panel displaying the directional rms error and the lower panel the amount of front-back confusions in percent. Chance level lays between the two dashed lines. As a reference, the *sim* condition as reported in Sec. III A. is shown as well. The average errors and standard deviations for all test conditions are shown in Table III.

Considering directional rms errors, the differences between BTE algorithms were not statistically significant. The amount of front-back confusions did not significantly differ between the *omni* and *NC* cases ($p \geq 0.27$). Due to the strong attenuation characteristics of the beamformer, the listeners could clearly identify sound coming from the back based on intensity cues in the *beam* condition. For this algorithm, some subjects verbally reported some front-back confusions, especially for sound being played at 0°, but responded correctly on the response map. For all scenes but the *office* scene, subjects performed significantly worse for the algorithms compared to the virtual simulations ($p \leq 0.02$). No statistical difference in directional errors between the algorithms and the simulation was found in the *office* case.

In Figs. 6 and 7, the results from positions played in the front and in the back were analyzed separately. As expected, for the *beam* condition, the rms error was lower in the front than in the back. This is due to the greater SNR in the front than in the back; improving therefore localization in the frontal area. Performance for all scenes but the cafeteria was similar for the beamformer as compared to the reference condition ($p \geq 0.25$).

For the *omni* and *NC* conditions, the error-rate was larger in the front than in the back, both in terms of rms errors and amount of front-back confusions. In the back, directional performance was similar to the reference condition. This similarity was significant only for the *cafeteria* scene ($p \geq 0.95$).

TABLE III. Mean results and standard deviations for the BTE conditions. The last column shows performance averaged across scenes.

	cafeteria			office			street			forest			all		
	omni	NC	beam	omni	NC	beam	omni	NC	beam	omni	NC	beam	omni	NC	beam
rms (°)															
mean	7.9	8.7	10.3	8.2	8.3	7.0	11.2	11.8	12.7	15.4	16.1	16.2	13.7	14.3	14.7
Std	4.0	3.9	4.0	3.9	3.7	4.7	3.1	3.2	3.2	2.8	4.2	2.8	2.4	2.4	2.0
f-b (%)															
mean	39.0	39.2	1.9	40.8	37.1	1.0	45.4	48.3	3.8	47.3	43.3	1.5	43.1	42.0	2.0
Std	11.2	8.9	3.2	12.2	11.9	2.0	7.9	8.3	5.5	6.4	10.1	3.1	5.8	6.0	2.5

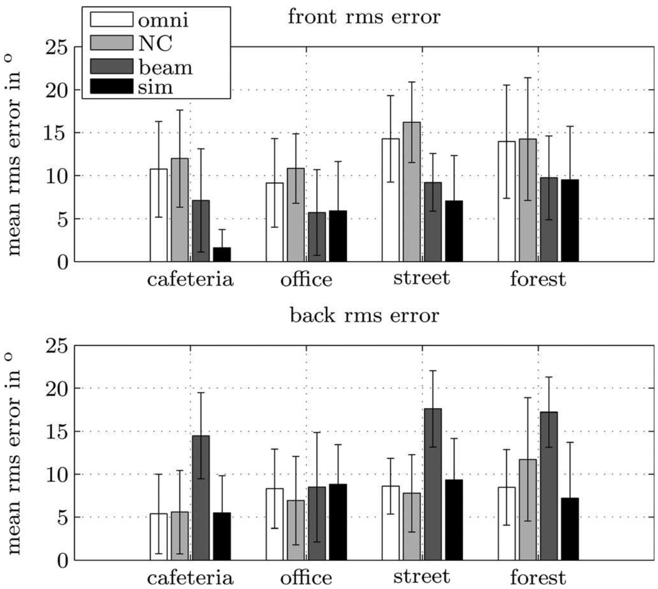


FIG. 6. Mean rms error for positions played at front ($|\theta| \leq 60^o$, above) and in the back ($|\theta| \geq 120^o$, below).

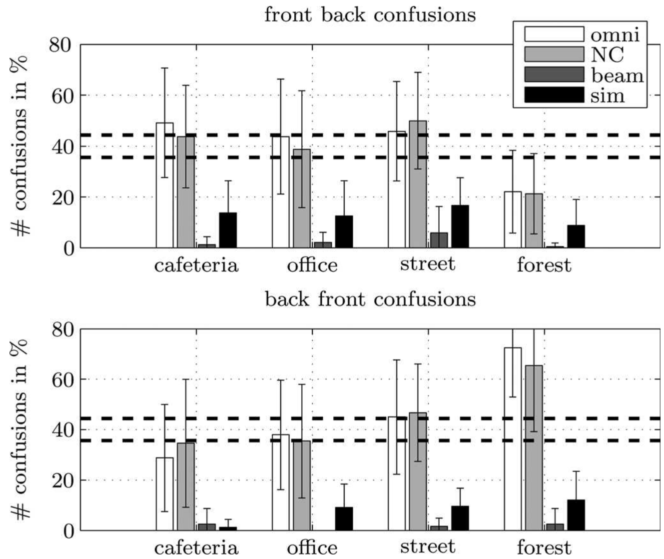


FIG. 7. Percentage of front-back confusions for positions played at front ($|\theta| \leq 60^o$, above) and in the back ($|\theta| \geq 120^o$, below). Note that the y-axis has been rescaled.

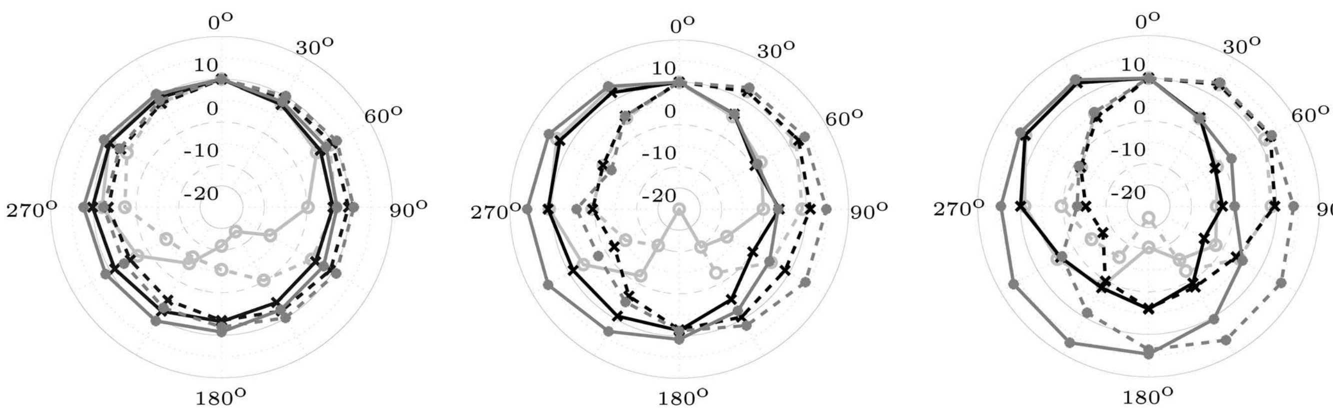
By separating the front-back confusions that occurred in the front from the ones in the back, a pattern emerges for the *forest* scene. It appears that the target signal was mostly localized in the front, whereas performance was close to chance for the other scenes. The reason for this can be explained by the spectral content of the target signal. It is the only signal that is essentially composed of frequencies above 2 kHz (see Fig. 1). In Fig. 8, the directivity patterns of the beamformer, the HRTFs measured at the entrance of the ear canal and at the position of the BTE microphones are represented. At low frequencies, the intensity diagrams for both HRTF measurement positions are similar. For the octave-band centered at 4 kHz, the effect of the pinna-loss is clearly visible with a difference of 10 dB. The BRTFs at this frequency band were similar for the front and the back. A sound composed of high frequency is therefore heard as coming from the front.

IV. DISCUSSION

The aim of this study was to investigate sound localization in realistic acoustical conditions in people with actual bilateral hearing aid algorithms. For this purpose, artificial environments with background noise have been reproduced using individual HRTF measurements and room simulations. The study involved various different aspects of human sound localization that are discussed separately in the following sections.

A. Sound localization in noise

In this study, the listeners had to localize a sound signal in diffuse background noise. It has earlier been shown (Lorenzi *et al.*, 1999; Langendijk *et al.*, 2001) that interfering noise has an impact on sound localization, depending on the intensity and the position of the noise relative to the target signal.



(a) $f_c = 250$ Hz (b) $f_c = 2000$ Hz (c) $f_c = 4000$ Hz

FIG. 8. Directivity characteristics of HRTFs measured at the ear canal with the open CIC microphones (black), behind the ear with the BTE microphones (gray) and of the beamformer (light gray) implemented at three different frequency bands. The directivities of the transfer functions measured at the left ear are plotted as a solid line. For the right ear, they are drawn in dashed lines.

In experiments I and II, the signals were played at 3 dB SNR and the SNR was further improved by the hearing aid algorithms. According to the findings of Lorenzi *et al.* (1999), sound localization is affected by noise at negative SNRs only. The improvement in SNR achieved by the algorithms had therefore no effect on localization performance, because the tested SNRs levels were above values at which localization performance is degraded. The difference in performance between localization with BTE hearing aids and the reference condition can solely be attributed to a distortion in spatial cues produced by the BTEs. The effect of noise on the amount of front-back confusions was however not addressed by Lorenzi *et al.* (1999).

In the two localization experiments discussed in this study, the rms errors of the test subjects were lower in the frontal hemisphere than in the back. This phenomenon appears in a series of sound localization experiments (Makous and Middlebrooks, 1990; Good and Gilkey, 1996; Gilkey and Anderson, 1995; Carlile *et al.*, 1997). In those studies, larger errors in the back than in the front for normal-hearing subjects were consistently observed for different stimuli (pulse trains, words, broadband noise) and attributed to the experimental setting of the tests. Carlile *et al.* (1997) and Makous and Middlebrooks (1990) evaluated localization performance using a head-pointing method, which required subjects to move their head to the direction of the sound being played. They assumed that the difference between front and back localization performance was due to the higher difficulty and time needed to move the head to the target in the back. This was different in experiments I and II, where subjects could report the relative position of the source directly on a screen in front of them, while the target was played continuously. It can be argued that the subjects had more difficulties in visualizing the exact positions of the loudspeaker in the back compared to the front, where direct visual feedback was available. This could have increased the uncertainty of their localization judgment and thus lower overall localization performance.

B. Localization of virtual sound sources

The combination of HRTFs and room simulations for the generation of virtual acoustical environments has been used in many past experiments. It is difficult, however, to compare the results of experiment I with earlier studies due to large differences in experimental settings. Nevertheless, in two studies (Hawley *et al.*, 1999; Rychtarikova *et al.*, 2009) similar acoustical conditions than in the *cafeteria* and *office* scenes were reproduced. In these two studies, the localization errors were of the same order than observed in experiment I, as discussed below.

In the study of Hawley *et al.* (1999), sound localization was evaluated for a target speaker along with interfering speech from the same talker. The number of competitors varied from none to three. The evaluation was carried out both using loudspeaker playback and virtual acoustics. They evaluated positions at the front only (-90^o to 90^o with steps of 30^o). Their experimental setup can be compared to the *cafeteria* condition of the present study. They measured a significant difference between real and virtual listening conditions but not between the number of competing talkers. Their rms errors were higher than in our *cafeteria* condition, being 10^o for real and 14^o for virtual playback (compared to 2.2^o to 3.9^o in experiment I). This difference can be explained by the small number of subjects doing the localization experiment (3) compared to this study (12). A single error on one trial results in an overall large increase in rms values. The average percentage of correct responses reported by Hawley *et al.* (1999) was of 96% and 83% for both reproduction methods against 95% and 86% in experiment I and are therefore similar.

More recently, Rychtarikova *et al.* (2009) investigated the localization of virtual sound sources in conditions similar to those of the *office* scene. Their study compared among others the localization of signals generated with loudspeakers versus sounds generated with HRTFs combined with room simulations ($T_{60} = 4$s). The sounds were reproduced using headphones and the HRTFs were recorded on an artificial head. In one of their setups, the target stimulus was a telephone signal. It was located either at 1 or 2.4 m of the listeners (1.5 in our study). In the reverberant room condition and for signals played at 1 m from the listeners, the average rms errors they obtained were 8.3^o and 9.1^o for loudspeaker playback and simulated BRIRs, respectively. For signals played at 2.4 m distance from the test subjects, the rms errors increased to 9.9^o and 11.5^o. In the anechoic room, performance improved to 7.3^o and 7.8^o. In this latter condition, the telephone signal was played at 1 m from the listener. In the *office* condition, the rms errors were 4.7^o and 7.2^o respectively for the *ls_open* and *sim* conditions. Those results are of the same order as in the anechoic settings in Rychtarikova *et al.* (2009), which implies that in experiment I reverberation was to small to have a significant effect on the rms error.

The system for auralizing the virtual scenarios applied static HRTFs and was therefore not able to cope with head movements. The differences in the number of front-back confusions between loudspeaker playback and simulation in the first experiment can be attributed to this effect. Although the test subjects were instructed not to move the head, unintentional head movements naturally may have occurred and could have been an advantage for the real versus virtual test conditions. This is especially true for the cafeteria scene, where the target signal had the largest low frequency content of all the scenes tested. This implies that interaural time differences are essential for the correct localization of the sound source. For positions played around 0^o (or 180^o) even small head movements can help finding the true position of the source.

C. Hearing aid localization

The bilateral hearing aid algorithms evaluated in this study had a significant impact on sound localization, although the differences in the average rms error between the omnidirectional and noise canceler conditions were rather low. This can be explained by the limited number of measured positions, which might have reduced the sensitivity of the experiment.

The main effect observed was an increase in front-back confusions caused by the loss of the pinna cues due to the positions of the microphones of the hearing aids. The directivity of the beamformer resolved these ambiguities. By analyzing separately the results of the front and back playback positions, it appears that the beamformer performed better in the frontal area than the other algorithms. It performed, however, much worse in the back due to reduced audibility of the target signal. In their study, Keidser *et al*. (2006) evaluated similar algorithms. Their reference, cardioid/cardioid and max. noise reduction conditions corresponded to *omni*, *beam* and *NC*, respectively. Their findings were consistent with the results of the second localization experiment, although the test conditions were different. The two first conditions were evaluated in quiet and the noise reduction algorithm was evaluated with a constant noise source at 80° with an SNR of 7 dB and the target stimulus was pink noise. They observed a slight but significant decrease in localization performance between the noise reduction algorithm and the reference condition. The cardioid microphone conditions also helped reduce front-back confusions.

Van den Bogaert *et al*. (2006) investigated sound localization with bilateral hearing aids in reverberant conditions ($T_{60} = 0.54$ s). The target stimuli were low-frequency and high-frequency noises and a telephone signal. The test subjects were normal hearing and impaired hearing subjects wearing real hearing aids. For the telephone signal, interfering noise was played at both sides of the subjects with a signal-to-noise ratio of 0 dB. The noise consisted of a multitalker babble. For the telephone signal, the normal-hearing subjects obtained an rms error of 11.8° in noise, which is higher than in the *office* condition (4.7°). The different test conditions between the two experiments could partly explain this difference. The spatial configuration of the noise sources between the two experiments differed. In Van den Bogaert *et al*. (2006), the noise source were played from two loudspeakers at both sides of the test subject whereas in our case the interfering noise was diffuse and played via 12 loudspeakers placed around the listener. For a fixed SNR, Langendijk *et al*. (2001) showed that sound localization was more difficult when the interfering noise and the target signal were close to each other. The local SNR was lower in Van den Bogaert *et al*. (2006) than in the present experiment. By looking at the results displayed in the localization plots showed in their study (Fig. 5), it appears that the addition of the masker increased the errors only at positions close to ±90°. In Van den Bogaert's study, hearing aids in their omnidirectional configuration were evaluated as well, altough only with hearing impaired listeners. They observed a degradation in sound localization performance (15.3° to 21.3°, respectively), which confirms the results found in experiment II.

In open hearing aid fittings, the acoustic wave bypasses the hearing aid and reaches the eardrum before hearing aid processing and playback. This direct acoustic path can provide intact localization cues to the hearing aid users and improve sound localization performance, provided enough residual hearing remains (Byrne *et al*., 1996). Furthermore, when the delay of the hearing aid is higher than 2 ms, the precedence effect ensures that the perceived position of the sound source is defined by the original acoustical wave (Litovsky *et al*., 1999). In the present experiments, this direct acoustic path has not been simulated as the focus was set on the effects on sound localization of the hearing aid algorithms only. For subsequent studies with hearing impaired listeners, this aspect must be considered so that the testing conditions are more realistic and closer to the hearing aid user daily experience.

V. CONCLUSION

In agreement with previous research, the outcomes of the localization experiments carried out in this study, suggest that by combining HRTFs with room simulations one can create acoustical environments that sound convincing and in which localization ability is preserved. A significant increase in front-back confusions with virtual playback was noticed only for one of the four scenes simulated. This is a common problem in virtual sound localization experiments and can be related to the inability of our sound reproduction system to cope with head movements. This could be improved by combining a head motion sensor with the system for virtual acoustics.

The localization experiments carried out in this study took place in noisy and realistic scenes in which hearing aids traditionally operate. The results are consistent with findings from earlier experiments that were carried out in the laboratory in much simpler acoustical conditions. In particular, the experiments presented here showed that bilateral hearing aids distort the spatial perception of sound. However, the algorithms tested represent only a small sample of what is available on the hearing aid market today. Specifically, new binaural algorithms that were designed to reproduce correctly the interaural cues have been developed. The real benefits of these algorithms need to be evaluated in realistic settings such as noisy and reverberant environments or multi-talker conditions. Moreover, other dimensions of spatial auditory perception such as the internalization, the perceived distance or the diffuseness of sound sources need to be investigated. The previously described setup allows the evaluation of these aspects.

The system for virtual acoustics is capable of reproducing environments that are more dynamic and closer to the real-world. In such environments, sound sources move along defined trajectories in space and in time. The behavior of adaptive algorithms is strongly linked to the environment in which they are used. Virtual acoustics could help to understand how spatial perception is affected by those algorithms and speed up the development of new binaural hearing aid prototypes.

ACKNOWLEDGMENTS

This project was funded by Grant No. 9646.1 of the Swiss Confederation's innovation promotion agency (CTI) in collaboration with Phonak AG. The authors are grateful to Juliane Schweizer and Peter Derleth for the many interesting discussions related to this work. The authors would further like to thank three anonymous reviewers for their very constructive comments on an earlier version of the manuscript.

Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., and Carlile, S. (**2010**). "A comparison of CIC and BTE hearing aids for three-dimensional localization of speech," Int. J. Audiol. **49**, 723–732.

Boymans, M., Goverts, S. T., Kramer, S. E., Festen, J. M., and Dreschler, W. A. (**2009**). "Candidacy for bilateral hearing aids: A retrospective multicenter study," J. Speech Language Hear. Res. **52**, 130–140.

Bronkhorst, A. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.

Byrne, D., Noble, W., and Glauerdt, B. (**1996**). "Effects of earmold type on ability to locate sounds when wearing hearing aids," Ear Hear. **17**, 218–228.

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196.

Duda, R., and Martens, W. (**1998**). "Range dependance of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Gardner, B., and Martin, K. (**1994**). "HRTF measurements of a kemar dummy-head microphone," MIT Media Lab Perceptual Computing Technical Report #280, pp. 1–7.

Gilkey, R., and Anderson, T. (**1995**). "The accuracy of absolute localization judgments for speech stimuli," J. Vestibular Res. **5**, 487–497.

Good, M., and Gilkey, R. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U., Puder, H., and Rass, U. (**2005**). "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," EURASIP J. Appl. Sig. Process. **18**, 2915–2929.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hershkowitz, R. M., and Durlach, N. I. (**1969**). "Interaural time and amplitude JNDs for a 500-Hz tone," J. Acoust. Soc. Am. **46**, 1464–1467.

Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (**2006**). "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal performance in hearing aid wearers," Int. J. Audiol. **45**, 563–579.

Kim, S., and Choi, W. (**2005**). "On the externalization of virtual sound images in headphone reproduction: A wiener filter approach," J. Acoust. Soc. Am. **117**, 3657–3665.

Köbler, S., and Rosenhall, U. (**2002**). "Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification," Int. J. Audiol. **41**, 395–400.

Langendijk, E. H. A., Kistler, D. J., and Wightman, F. L. (**2001**). "Sound localization in the presence of one or two distracters," J. Acoust. Soc. Am. **109**, 2123–2134.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Makous, J., and Middlebrooks, J. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2186–2200.

Matsumoto, M., Yamanaka, S., Tohyama, M., and Nomura, H. (**2004**). "Effect of arrival time correction on the accuracy of binaural room impulse response interpolation. Interpolation of room impulse responses," J. Audio Eng. Soc. **52**, 56–61.

Moeller, H. (**1992**). "Fundamentals of binaural technology," Appl. Acoust. **36**, 171–218.

Noble, W., and Byrne, D. (**1990**). "A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical plane," Br. J. Audiol. **24**, 335–346.

Noble, W., and Gatehouse, S. (**2006**). "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ)," Int. J. Audiol. **45**, 172–181.

Rife, D., and Vanderkooy, J. (**1989**). "Transfer-function measurements with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444.

Rychtarikova, M., van den Bogaert, T., Vermeir, G., and Wouters, J. (**2009**). "Binaural sound source localization in real and virtual rooms," J. Audio Eng. Soc. **57**, 205–220.

Schimmel, S., Mueller, M., and Dillier, N. (**2009**). "A fast and accurate shoebox room acoustics simulator," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 241–244, available at roomsim.sourceforge.net (Last viewed December 27, 2010).

Van den Bogaert, T., Carette, E., and Wouters, J. (**2011**). "Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna," Int. J. Audiol. **50**, 164–176.

Van den Bogaert, T., Klasen, T., Moonen, M., Deun, L. V., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: Without is better than with," J. Acoust. Soc. Am. **116**, 515–526.

Wallach, H. (**1940**). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368.

Wightman, F., and Kistler, D. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F., and Kistler, D. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853.